APPENDIX A

SAMPLE SELECTION

Appendix A: Sample Selection for Quality Review

This section explains in detail the procedures for selecting the samples for nonmonetary determinations and lower authority appeals review. Information is provided regarding the requirements of a sampling system, and options available for selecting samples. Flexibility of **options** has been included in the procedures explained. Each State should select the option best suited to their particular operation. The option preferred by the N.O. will be indicated with reasons for the recommendation.

WHAT DOES SAMPLING REQUIRE?

The sampling methodology for nonmonetary determinations and lower authority appeals contains five distinct steps.

1) Identify, find, or gather data elements for sampling Universe files;

2) Extract or collect data to create the Universe files;

3) Determine which transactions to select for the sample;

4) Select the cases to review;

5) Create output reports and files of the selected cases.

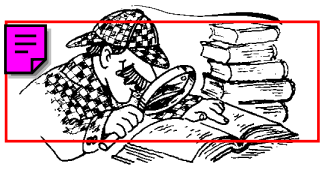
How these five steps are accomplished is the SESA's choice. Not all States have the same level of automation, and varying file structures may lend themselves to different sampling approaches.

The SESA may use an automated or manual method to create the universe files and select the samples. Throughout this section of the Appendix, both methods will be discussed. Various details and tips on how to decide which is the best approach for your State will be provided.

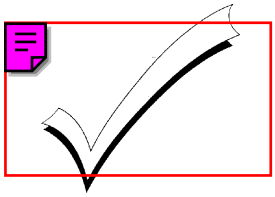
Appendix A: Sample Selection for Quality Review

Step 1. - Gather Data for the Sampling Universe

Collect Required Data



The first step is to gather or have access to **THE UNIVERSE (ALL** of the particular transactions to be reviewed). **It is essential that every transaction or item meeting the criteria be included.** This means that all possible sources or locations of the transactions must be searched. For example, since nonmonetary determinations generated in various units such as BPC and BQC are included by definition, check to be certain they will all be included in the universe.



Be sure to check that only valid transactions are included.

For instance, nonmonetary determinations which are generated for the sole purpose of establishing an overpayment amount are not valid in terms of the definitions.

Refer to the definitions in the UI Reports Handbook No. 401.

Depending upon the automation level of the SESA, the creation of the universe may require the extraction of data from a computer (mainframe or PC) or the keeping of a manual list or searching file drawers.

If the automated sampling method is chosen, the SESA ADP staff is responsible for creating universe files which contain the requested information. The National Office has developed specifications of the minimum data needed for each sampling universe. The SESA ADP staff is responsible for creating the programs and/or utilities to extract or gather the requested data elements.

Each sample being reviewed will be selected from a universe which includes the nonmonetary determinations or lower authority appeals which were dated within the three

months in the quarter preceding, which is referred to as the review quarter.

Once the elements for the universe files have been identified, and the extraction program created, it is recommended that the BTQ reviewer examine a small cross section of the records to verify that the data elements are correct and the proper time frames are being followed. **It is strongly recommended that the data verification of the universe files be accomplished before the actual time period that the data is needed and the actual universe files are created.**

After the BTQ reviewer has approved the data elements and time frame, the SESA ADP staff should establish procedures for building the universe files, selecting the sample cases and saving the universe files.

Schedule Data Capture

We strongly recommend that in building the universe files nonmonetary determinations and lower authority appeals be captured as they occur.

This is important since the desired transaction may be superseded by an subsequent transaction and the desired information is no longer readily available. This can be especially true in highly automated States where data fields are often overlaid with the most recent information.

Some States may be able to reconstruct events by using daily transaction logs maintained in their data processing environment. For the most part it is still better to capture the transactions as they occur during the time period to be reviewed.

In manual record storage systems, the transaction source information may be filed in a filing cabinet or logged by docket number. The universe to be sampled should be

isolated in some manner so that the integrity of the data will not be compromised when it is time to select the sample.

Step 2. - Collect Data to Create the Universe

Data Collection Automated Approach

Each SESA has the option to collect the data for the universe files any way they choose. If the necessary file is automated, the SESA ADP staff should create a program to gather the data. This program can be written using any computer language or utility such as Easytrieve, FileAid, or SAS. The N.O. recommends that the SESA follow the N.O. file layouts provided. The resulting transaction files can then be used as input into the N.O. Sample Determination program.

Data Collection Manual Approach

For files that are not automated, the SESA must collect the information manually. This can be accomplished many ways but a simple method can be the creation of a paper list.

For example: as lower authority appeals are completed, each hearings officer will write the docket number and the date the appeal was heard on a paper list. Throughout the quarter, the list is updated possibly onto numerous sheets of paper. At the end of the quarter, all the lists are assembled as the universe and the number of completed appeals are sequentially numbered and counted.

Save Universe Files

It is important to save the original universe files.
States **MUST** save the universe files from which samples are selected for data validation purposes.

Step 3. - Determine Records for the Sample

Perform Calculations

The third step is to determine which records to select for the sample. The formulas used to determine which records to

select must be the formulas provided by the National Office.

To perform the calculations, three numbers are needed:

- **Total Records in the Universe** - Once the universe has been created, a count of all the transactions in the universe must be performed. This count is represented by "P" in the calculations.

- **Number of Records to Sample - NONMONETARY DETERMINATIONS.** The number of cases to sample for nonmonetary determinations depends on the **total** number of nonmonetary determinations reported by the State in the preceding calendar year. States reporting 100,000 or more nonmonetary determinations will sample 50 separation issues and 50 nonseparation issues each quarter for quality review. States reporting fewer than 100,000 determinations in the preceding calendar year will sample 30 separation issues and 30 nonseparation issues each quarter for quality review. States may sample larger numbers if they choose, but all of the determinations sampled must be reviewed and entered into the database in order to preserve the validity of the sample. Before running the sampling routine, the universe of nonmonetary determinations should be sorted by separation issues and nonseparation issues so that an independent sample can be drawn from each.

- **Number of Records to Sample - LOWER AUTHORITY APPEALS.** The number of cases to sample for lower authority appeals quality depends on the **total** number of lower authority appeals decisions reported by the State in the preceding calendar year. States reporting 40,000 or more decisions will sample 40 decisions each quarter for quality review. States reporting fewer than 40,000 decisions in the preceding calendar year will sample 20 decisions each quarter for quality review. States may sample larger numbers if they choose, but all of the determinations sampled must be reviewed and entered into the database in order to preserve the validity of the sample.

  NOTE: Interval sampling can be implemented with no difficulty in States whose automated systems allow appeals disposed of by withdrawal, dismissal, or "no show" to be identified by code and excluded from the universe. If these types of disposals are indistinguishable and cannot be excluded from the universe before sampling, there are three alternatives to pursue.

  The first, and possibility easiest, method would be to add a marker with which to indicate those lower authority appeals in which an actual hearing led to a

decision. The marked appeals would constitute the universe for pulling a random sample by interval sampling.

As a second alternative, ADP could be asked to randomize the lower authority appeals file transactions within the quarter being reviewed. Randomizing routines are readily available, and once the file is randomized, the reviewers can begin drawing their sample from the top of the list, examining each record to be certain it meets criteria before saving it to the sample list. As many records as necessary can be pulled and examined until the required number is obtained, and the result will be a random sample of valid appeals decisions.

A third approach requires running the sampling routine three times. First, draw a random sample of twenty decisions from the universe, removing them. Then draw a second sample of 5 decisions from the universe remaining, removing them. And finally, draw a third sample of 3 decisions from the universe remaining. If all twenty of the first sample are valid hearings decisions, the next two samples will not be needed. If only 15 of the first sample are valid, move to the second sample of 5. If all of them are valid, stop. But if only 4 of them are valid, the third sample will be needed. The drawback here is that to preserve sample integrity, all of the valid cases from the third sample must be reviewed and entered, even if only one is needed.

- **Random Number** - This is the third critical number necessary to perform the sample calculations. It is represented as "R" in the formulas. Random numbers may be obtained from any statistics manual.

FORMULAS TO IDENTIFY RECORDS FOR QUALITY SAMPLES

The following instructions apply to the steps needed to determine which records to select for the sample. These steps must be repeated for each sample that will be selected.

Whether the SESA chooses to perform these calculations manually, develop an automated method of performing these calculations, or use the N.O. supplied software, these formulas must be followed.

- **A count of the total number of transactions in the universe must be performed**. If the universe is created manually, the BTQ reviewer must count the transactions. If the creation of the universe is automated, the SESA ADP staff can supply this number.

This number is represented by "P" in the calculations. Note that for nonmonetary determinations, a sort must be performed to divide the universe into its components of separation issues and nonseparation issues before proceeding.

- **Determine the number of cases to sample**. Based on the number of nonmonetary determinations and lower authority appeal decisions reported by the State in the previous calendar year, determine the the number of cases to sample for each. The letter "N" represents sample size in the calculations.

- **Obtain a Random Number**. Refer to a list of random numbers supplied in any statistics handbook. In the calculations, "R" represents the random number.

After the above mentioned numbers are identified, several calculations must be performed.

CALCULATIONS

First, determine the sampling interval (K). This is accomplished by dividing the sample size into the universe size. If the result of this calculation is not a whole number, round the result to the nearest interger.

K = P/N (round to the nearest integer)

Second, determine the starting point (I) within the universe. This is accomplished by multiplying the sampling interval (k) by the random number (R) and adding .5. The result of this calculation should be truncated to the nearest interval.

I = (R*K) + .5 (truncate to the nearest integer)

Next, "N" cases must be selected. This is accomplished by selecting pairs of cases (J) until all the cases have been identified. First, the number of pairs must be determined by:

If N is even, J = 0, 1, 2, ... (1/2N - 1)
If N is odd, J = 0, 1, 2, ... 1/2(N - 1) - 1, the remaining case is calculated separately.

Once the number of pairs is determined, the cases are selected by using the following formulas:

I + JK and (P - JK) - I + 1

The remaining (odd) case is calculated by:

I + 1/2(N - 1)K

CALCULATIONS

For example if: P = 43, N = 5, R = .261

K = 43/5 = 8.6 = 9
I = (.261 * 9) + .5 = (2.349) + .5 = 2.849 = 2

Since N = 5,
J = 1/2(5 - 1) - 1 = 1

The following records would be selected:

	I + JK	(P - JK) - I + 1
when J = 1	2 + (1*9) = 11	(43 - 1*9) - 2 + 1 = 33
when J = 0	2 + (0*9) = 2	(43 - 0*9) - 2 + 1 = 42

the remaining case is calculated by: I + 1/2(N - 1)K

2 + 1/2(5 - 1)9 = 20

The first record selected is 11, the second is 33, the third record selected is 2, the fourth is 42, and the fifth record is 20.

If: P = 244, N = 10, R = .743

K = 244/10 = 24.4 = 24
I = (.743 * 24) + .5 = (17.832) + .5 = 18.332 = 18

The following records would be selected:

18, 227, 42, 203, 66, 179, 90, 155, 114, and 131.

National Office System to Identify the Sample

The N.O. has created software to perform the calculation step of the sampling process. If a SESA chooses to use the N.O. software (COBOL) to draw the sample, the three critical numbers mentioned above must be provided as input before using the software.

Once the software has been run, the records for sampling will be identified and the software will produce a report listing the record numbers selected.

Other Automated Approach (e.g. Easytrieve)

The SESA may choose to use another automated method of identifying which records will constitute the sample. However, it is imperative that the formulas described on page A-8 be used to ensure that the sample selection is non-biased. ***If a SESA feels that there may be a better method for random sampling, they must submit the proposed approach*** *in writing to their Regional Office and receive written approval before proceeding.*

Manual Sample Identification

The SESA may choose to perform the sample identification manually. The steps to perform the identification are no different than the automated process. The formulas on page A - 8 will need to be calculated manually and the three numbers: total records in the universe (P), the number of records to sample (N), and the random number (R); will also be needed to perform these calculations.

Calculate the record numbers using the supplied formulas. Once the record numbers are calculated, a list of these numbers should be created. This list will be used to identify the universe records to select. The selection of these records is explained in the next section.

Steps 4 & 5 - Select Cases and Create Sample Files

Appendix A: Sample Selection for Quality Review

The last step of the sampling process involves the creation of files containing the selected sample cases. This step uses the universe file from step two and the calculated record numbers from step three to create the sample file. The calculated record numbers from step three identify which records from the universe file will comprise this file.

The sample case file must contain the skeleton fields for nonmonetary determinations and lower authority appeals identified in Attachment A to UIPL 10-96, Benefits Timelapse and Quality Measures, plus the claimant's social security number.

Manual Reports

As stated earlier, **the minimum amount of information needed for the sample file is the claimant's social security number, plus the data for the "skeleton fields" identified in Appendix A to UIPL 10-96.**

If a SESA used a paper list to assemble the universe, the records on the paper list will be sequentially numbered. The sample is selected by matching this sequential number to the case numbers determined in the calculations. The records that matched would be used to create a list. This list could be created by:

- highlighting the records on the original list;
- creating a separate list of just the selected records; or
- physically pulling the files from their storage area.

National Office Sampling System

As stated earlier, States are responsible for creating the programs/utilities necessary to extract the data elements for the universe files for each Sample. The National Office Sampling System consists of two COBOL programs, which are fully documented in ET Handbook 407, Appendix A. The first program, PICKNMBR, will calculate which record numbers need to be sampled and the second program, SAMPS0nn, was designed to perform

the sampling selection process for Revenue Quality Control (RQC). The RQC COBOL routine which most closely resembles the routine necessary to draw the nonmonetary determinations quality sample is the Status Determination (see page A - III - b - 9 of the Appendix referred to above) but it would need to be adapted for use in sampling nonmonetary determinations and lower authority appeals.

Appendix A: Sample Selection for Quality Review

How Does This Software Work?

A control record must be created in order to execute the Sample Determination program, PICKNMBR. The SESA ADP staff must supply most of the data needed on this file. The National Office program, PICKNMBR. will read this file, calculate which records to select for the sample, and write the record numbers to a file.

PICKNMBR

This program should be used for every sample process (both manual and automated). The calculations performed by this program were designed to ensure a non-biased systematic sample. This program is designed to be used as a stand-alone program and is not dependant upon the method that will be used to select the sample cases.

This program can be used if the Universe is:

- not stored using the N.O. format;
- a paper list;
- kept as transactions occur; or if
- kept in a filing cabinet.

Detailed descriptions of PICKNMBR and of the National Office COBOL sampling programs are provided in ET Handbook 407, Appendix A.

PICKNMBR Processing

The narrative below describes the processing steps that are performed in the PICKNMBR program. These processing steps are also illustrated in flowchart format.

0000-DRIVER-ROUTINE.

This section is the main routine for the program. This routine calls all of the other routines.

0010-LIST-HEADING
0020-LIST-HEADING

These sections control printing of the report page and column header information, line count, and page advancement.

0011-CS011, 0031-CS031, 0041-CS041, 0042-CS042, 0043-CS043, 0051-CS051, 0061-CS061.

These sections identify either nonmonetary determinations or lower authority appeals and corresponding year/quarter fields for the activity being processed.

0100-OPEN-ROUTINE.

This section opens the input file CNTRL-DATA, and output files SELECT-NUMBERS, PICKNUM-LIST and reads the CNTRL-DATA file.

0110-CNTL-OPTION.

This section determines which function is being processed.

PICKNMBR Processing continued

0120-CNTL-ERROR.

This section validates the three CNTRL-DATA file fields CNTRL-RANDOM-ALF, TRANS-REC-CNTRL-ALF, and SAMPLED-NMBR for non-numeric values. These fields must be numeric for the program to execute. To assist in the validation, a STOP-FLAG field is incremented by a certain amount. As a result, a comparison is made between the incremented STOP-FLAG and a value in the range 0 through 7. Within this range, different error messages will be displayed depending upon the error detected, then the program is terminated.

0130-FIPS-TABLE.

This section searches SESA-ID in the FIPS table to find the exact state name associated with its abbreviation.

0140-SPL-TABLE.

This section searches the SAMPLE-TYPE field of the CNTRL-DATA file for a corresponding match in the sample table (SPL-TYPE-DATA). If a match occurs, the sample type abbreviation is replaced by the exact sample type description to be utilized in the output report formats.

0200-CALC-SKIP-INTERVAL.

This section calculates the SKIP-INTERVAL (K) utilizing the following equation K=P/N. P: the total number of records, N: sample size.

PICKNMBR Processing continued

0300-INITIAL-CASE.

This section calculates the initial sample case number (I). It is determined by truncating the result of I = R*K + 0.5. The INITIAL-CASE field (I) is defined as a 5-position numeric integer. The right side of the equation I = R*K + 0.5 yields a real number, thus allowing (I) to truncate the result of the calculation. R: CNTRL-RANDOM-ALF, and K: SKIP-INTERVAL.

0310-CHECK-ODD-EVEN.

This section determines whether the number of records (N) to be selected for the sample (SAMPLED-NMBR) is either odd or even. If (N) is odd, the 0320-ODD-RTN procedure is executed.

0320-ODD-RTN.

This routine calculates the additional number that was described in the random function formula. The equation is as follows: ONE-MORE-REC = I + 1/2(N-1) * K.

0330-CREATE-REC.

This routine writes the calculated record numbers to an output data file (SELECT-NUMBERS) and an output print file (PICKNUM-LIST). A record counter (MATCH-CNTR) is incremented by one each time a record is added to the files.

PICKNMBR Processing continued

0400-REMAINING-NUMBER.

This section performs the calculations to determine the remaining numbers for the sample. As the balanced systematic sample of the random function formula, if N is even, N/2 pairs of records are selected. If N is odd, the iteration number is as follows: 0, 1, 2, ...1/2(N-1) - 1 and 1/2(N-1) - 1 pairs of records are selected. See Exhibit A-3 for the actual formulas needed to determine the records to select.

0500-CALC-SKIPINTERVAL.

This section will calculate the SKIP-INTERVAL-B for a transaction sample size less than 200 using the equation K=P/N (K not rounded).

0600-CALC-INITIAL-CASE.

This section calculates the initial sample case number(I) for the transaction sample size less than 200. It is determined by truncating the result of I = R*K + 0.5.

0700-SELECTED-NUMBERS.

This section performs the calculations to determine the numbers to select for the sample. This procedure is based on the transaction sample size less than 200. The second record is calculated by adding the skip interval (not rounded) to the initial case (truncated) and rounding the result. The remaining numbers are calculated by adding the skip interval to the previous (not rounded) number and then rounding the result. This process is continued until all of the records have been calculated.

PICKNMBR Processing continued

0800-CHK-SPL-NBR.

This section verifies that the number of records written to the SELECT-NUMBERS files equals the number of records to be sampled (SAMPLED-NMBR) in the CNTRL-DATA file. If these fields are not equal, and error message is displayed.

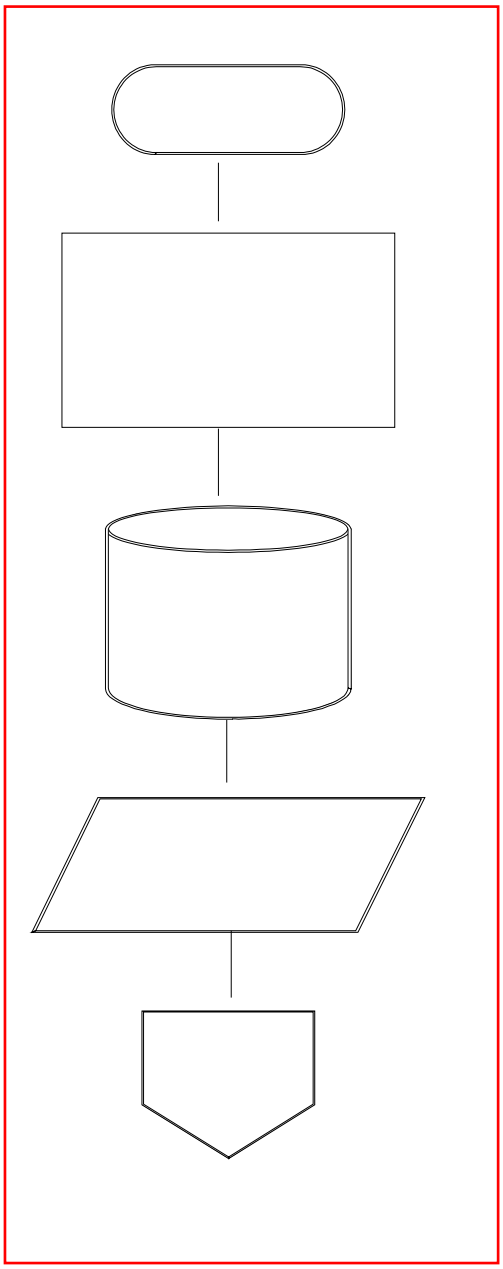
0900-TRAILER-LIST.

This section prints the information that was used to perform the calculations and select the record numbers.

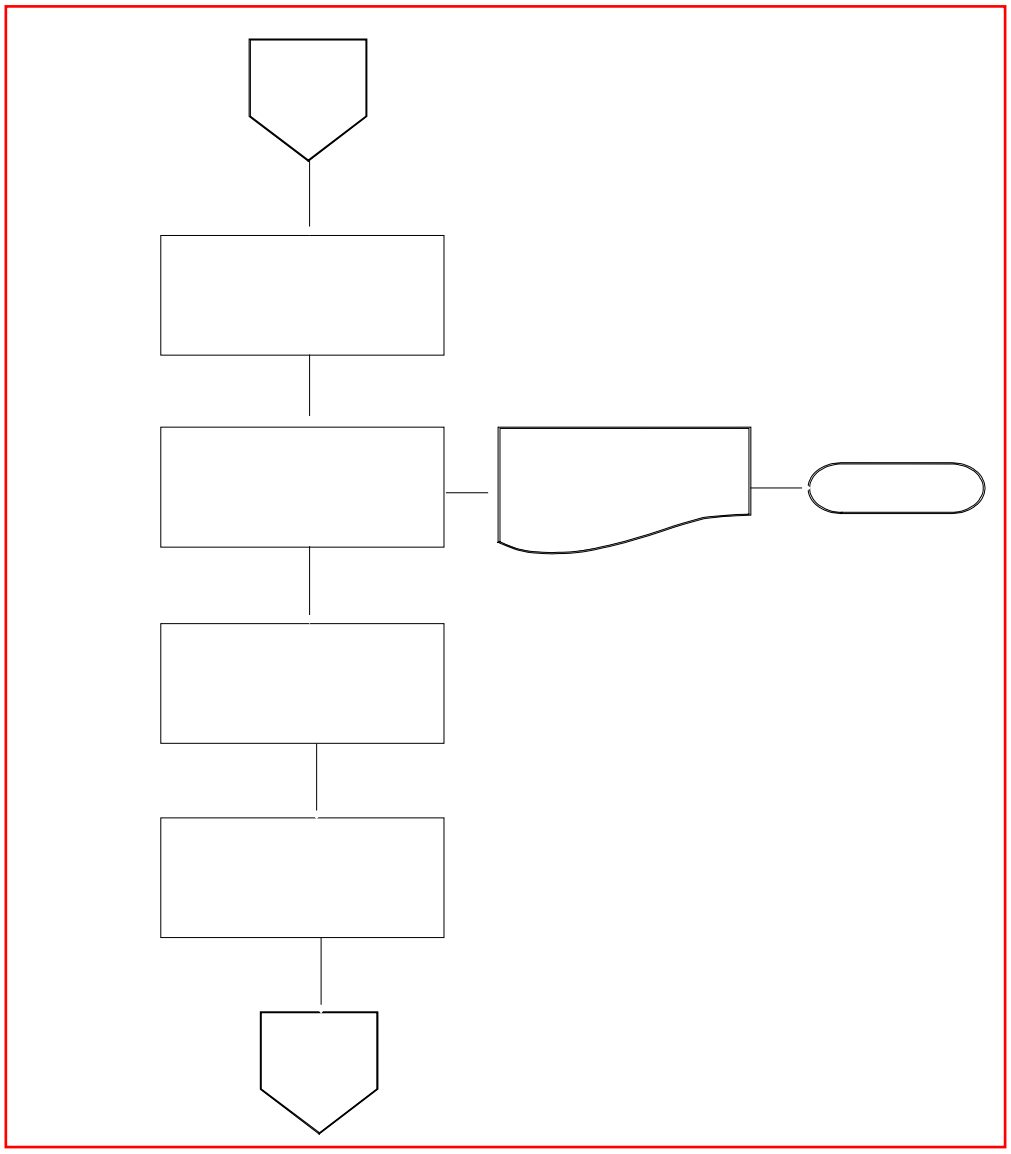
9999-CLOSE-FILE.

This section closes the files CNTRL-DATA, SELECTED-NUMBERS, and PICKNUM-LIST.

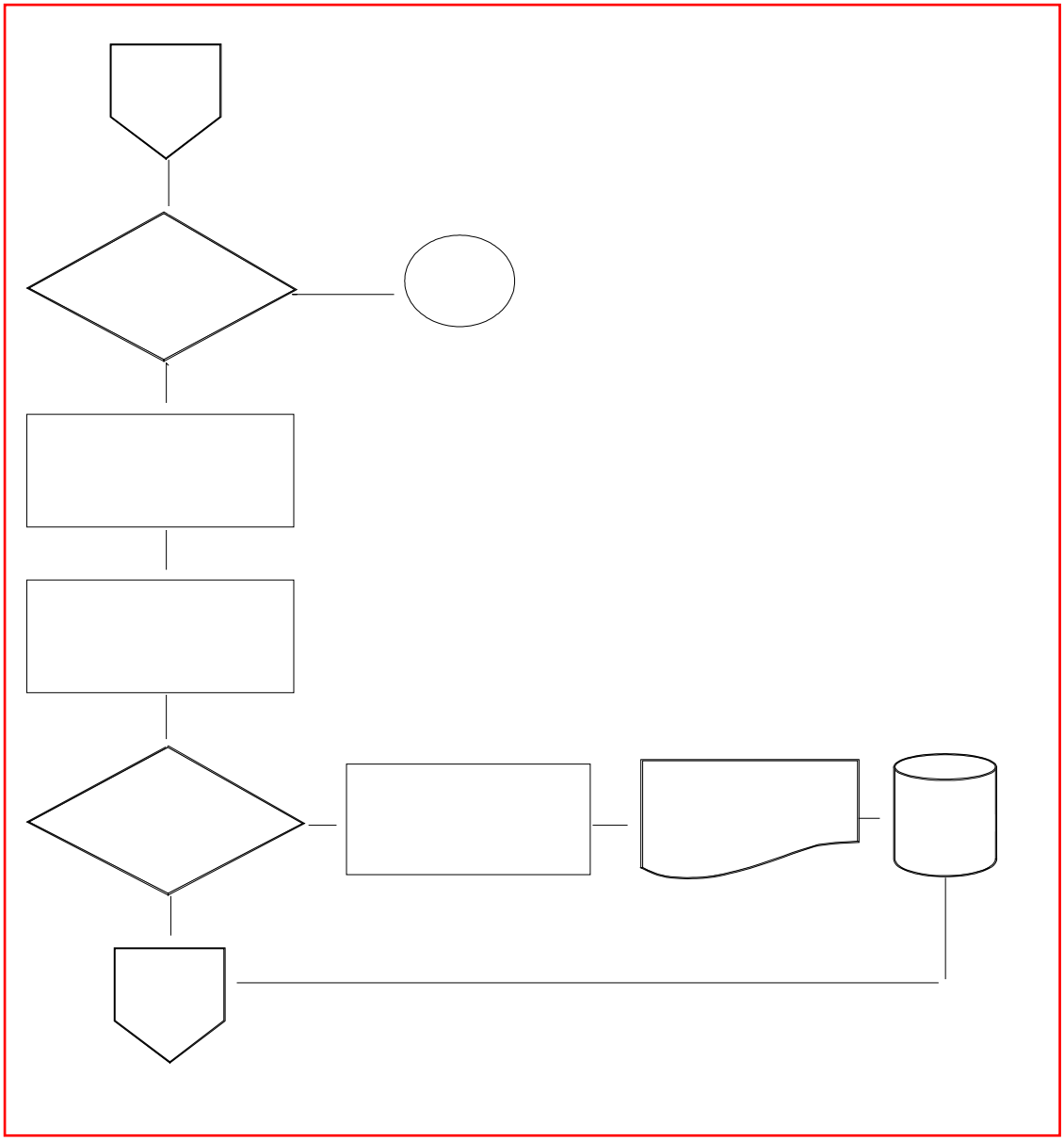
PICKNMBR



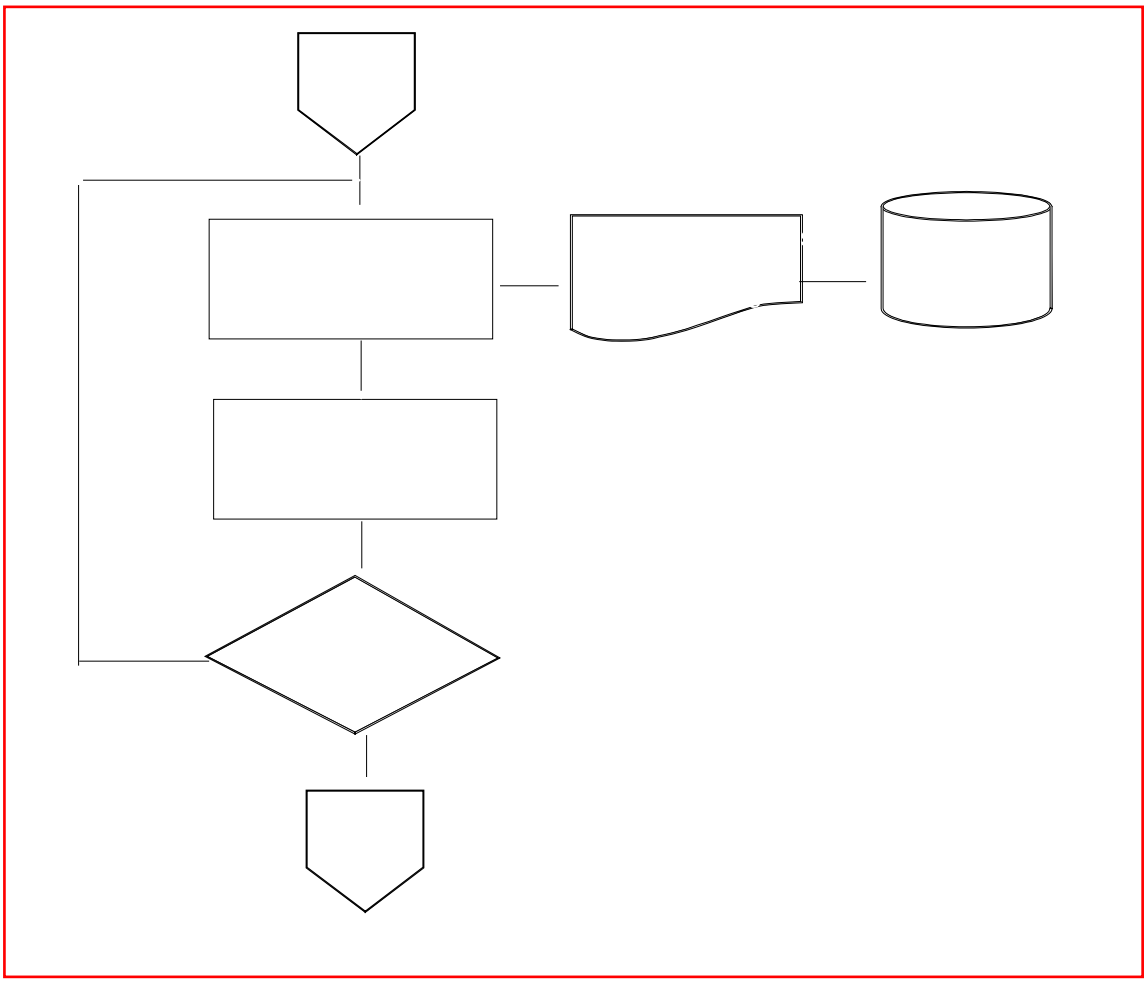
PICKNMBR



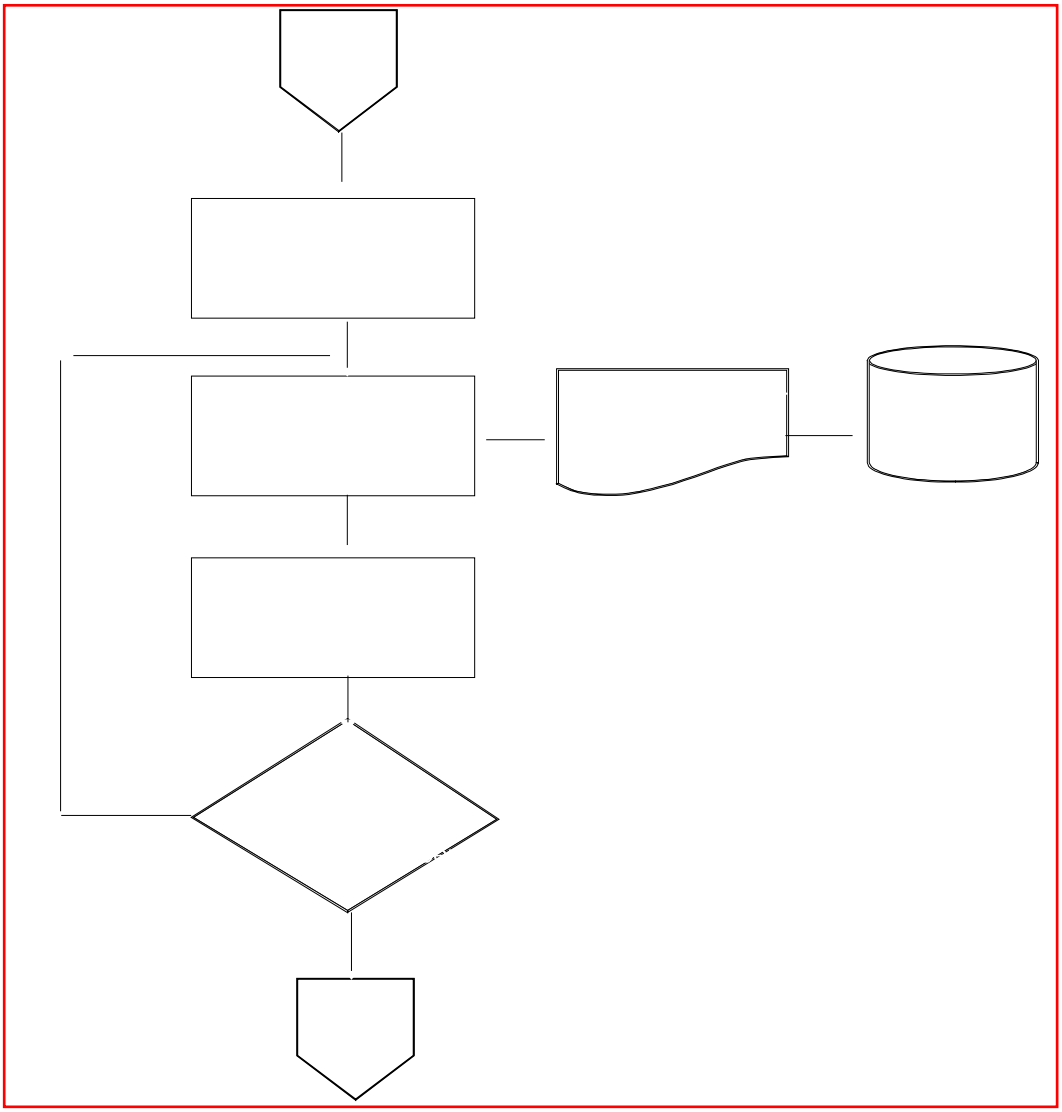
PICKNMBR



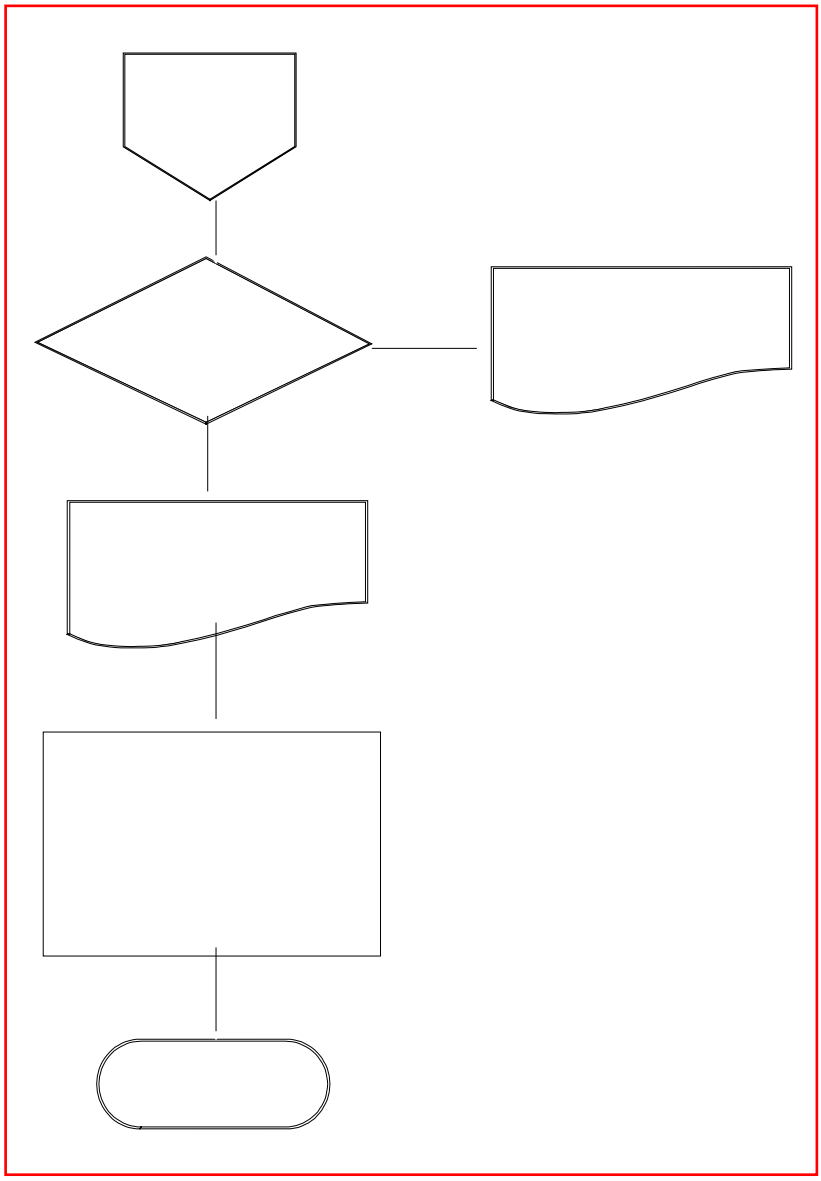
PICKNMBR



PICKNMBR



PICKNMBR



PICKNMBR